

## PENERAPAN ALGORITMA AHC ALGORITHM DALAM APLIKASI PEMBAGIAN KELAS SISWA BARU

Marjiyono<sup>1)</sup>

<sup>1)</sup> Teknik Informatika STMIK AMIKOM Yogyakarta  
Jl Ring road Utara, Condongcatur, Sleman, Yogyakarta 55281  
Email : aji.marjiyono@gmail.com<sup>1)</sup>

### Abstrak

*Pengelompokan siswa sering dilakukan di sekolah. Dengan mengelompokkan siswa dengan baik maka diharapkan akan mempermudah dalam proses belajar mengajar.*

*Agglomerative Hierarchical Clustering Algorithm atau biasa disebut AHC Algorithm merupakan suatu algoritma untuk mengcluster berdasarkan jarak kedekatan obyek sehingga metode ini dapat diterapkan dalam proses pengelompokan siswa.*

*Untuk dapat menerapkan algoritma ini maka dibuat suatu aplikasi berbasis web yang dapat mengcluster siswa dengan baik. pembuatan aplikasi dilakukan dengan perencanaan, perancangan dan implementasi. Dalam implementasi, digunakan array multidimensi untuk menyimpan data sementara pada perulangan perhitungan. Setelah aplikasi selesai dibuat, maka di bandingkan dengan perhitungan manual sehingga dapat di ketahui bahwa algoritma di dalam aplikasi telah berjalan dengan baik.*

**Kata kunci:** clustering, siswa, AHC Algoritm, aplikasi.

### 1. Pendahuluan

Pengelompokan siswa perlu dilakukan untuk mempermudah dalam proses belajar-mengajar. Dengan dikelompokkannya siswa berdasarkan kesamaan kemampuan akademik maka guru dapat menyesuaikan cara dalam memberi pelajaran di setiap kelompok atau kelas. Sekolah dengan kelas yang heterogen adalah sekolah yang memungkinkan siswa untuk mencapai hasil yang positif dalam pembelajaran, baik siswa yang prestasinya tinggi atau rendah[1]. Untuk dapat menerapkan hal ini diperlukan cara yang tepat agar hasil pengelompokan dapat maksimal dan mudah dilakukan.

Salah satu cara yang dapat digunakan untuk mengelompokkan siswa adalah dengan clustering. Clustering adalah proses pengelompokan kumpulan dari objek-objek yang anggotanya memiliki kesamaan menjadi class-class dari objek yang sama. Tujuan clustering adalah mengelompokkan sekumpulan data atau

objek yang memiliki kesamaan dalam satu cluster dan berbeda dengan objek pada cluster lain [2].

Beberapa penelitian tentang clustering telah dilakukan, salah satunya adalah oleh Margo Ridho Leksono dan Ari Eka Prastiyanto dengan judul “penerapan algoritma clustering untuk mengelompokkan ketertarikan siswa berdasarkan aktivitas di model pembelajaran elektronik” mereka mengcluster siswa berdasarkan ketertarikan dari model pembelajaran[3].

Penelitian tentang clustering juga dilakukan oleh Humasak T.A. Simanjuntak dengan judul “implementasi metode hierarchical clustering pada data genetic mikroarray” pada penelitian tersebut dilakukan clustering dengan menghitung jarak antar gen[4].

Clustering merupakan pengelompokan data objek ke dalam class yang memiliki kesamaan. Kesamaan (similarity) diantara objek-objek ditaksir menurut Distance Measure. Hasil dari clustering adalah cluster, yang memiliki sekumpulan objek dengan kesamaan tertentu dalam satu clustering dan berbeda dengan objek pada cluster lain. Aktivitas untuk melakukan clustering disebut Clusteringanalysis [2].

Clustering termasuk unsupervised classification (tidak mengetahui label dari class dan jumlah class yang akandibentuk). Salah satu penggunaan clustering adalah pada data genetik mikroarray [2].

Metoda clustering dapat diklasifikasikan atas 5 kategori, antara lain, Partitioning Method, Hierarchical Method, Density-based Method, Grid-based Method, Model-based Method[2]. Yang menjadi fokus pada kajian ini adalah Hierarchical Method.

Hierarchical Clustering Method (HCE) merupakan salah satu metoda yang digunakan dalam melakukan clustering[2]. Clustering-clustering yang dibentuk dari HCE berupa sebuah hirarki (tree). Pada hierarchial clustering, data tidak dibagi menjadi sebuah clustering dalam satu langkah tetapi dalam serangkaian clustering. Pembagian dapat dilakukan dari sebuah clustering yang berisi semua objek ke beberapa clustering dalam jumlah n dimana masing-masing mengandung sebuah objek [2].

Kualitas dari *hierarchical clustering method* dipengaruhi oleh kemampuan metode ini untuk membuat keputusan yang tepat dan sesuai ketikamenggabung atau memisah *clustering*. Ketika keputusan untuk menggabung atau memisah telah dilakukan, maka keputusan tersebut tidak dapat dirubah lagi.

Metoda *Hierarchical clustering* diklasifikasikan dalam dua jenis yaitu *Agglomerative* dan *Divisible Hierarchical Clustering* tergantung pada komposisi hirarkis dilakukan dari bentuk bottom-up atau top-up[3]. Metode yang digunakan pada kajian ini adalah Metoda *agglomerative Hierarchical Clustering*. Metode ini dilakukan dengan serangkaian pengelompokan objek (banyak objek menjadi satu bagian). Dimulai dengan menempatkan setiap objek pada *clustering* dan kemudian menggabungkan struktur *clustering* tersebut menjadi *clustering* yang lebih besar sampai semua objek berada pada satu *clustering* yang sama atau sampai memenuhi kondisi yang telah ditentukan. Algoritma ini dilakukan dengan pendekatan bottom-up [2].

Algoritma yang digunakan pada *Hierarchical Clustering Method* adalah *Hierarchical Clustering Algorithm*. Pada algoritma tersebut, jika diberikan sejumlah N objek yang akan di-*clustering* dengan proximity matrix N\*N, maka [3]:

- Diawali dengan membuat setiap objek menjadi satu *clustering*. Jika tersedia sejumlah N objek yang akan di-*clustering*, maka akan terbentuk *clustering* sebanyak N yang berisi hanya satu objek.
- Cari pasangan *clustering* yang terdekat dan digabungkan menjadi satu *clustering*. *Clusteringclustering* dapat digabungkan menjadi satu *clustering* jika memiliki objek dengan kesamaan terdekat terhadap objek pada *clustering* lain.
- Hitung *distance* atau kesamaan antara *clustering* yang baru dibentuk dan *clustering-clustering* lain yang sudah ada sebelumnya.
- Ulangi langkah b dan c sampai semua objek yang di-*clustering* berada pada satu *clustering*.

*Agglomerative Hierarchical Clustering* merupakan metode dengan menggunakan strategi disain *Bottom-Up* yang dimulai dengan meletakkan setiap obyek sebagai sebuah *cluster* tersendiri (*atomic cluster*) dan selanjutnya menggabungkan *atomic cluster* – *atomic cluster* tersebut menjadi *cluster* yang lebih besar dan lebih besar lagi sampai akhirnya semua obyek menyatu dalam sebuah *cluster* atau proses dapat pula berhenti jika telah mencapai batasan kondisi tertentu [2].

Adapun ukuran jarak yang digunakan untuk menggabungkan dua buah obyek cluster adalah *Ecluidean Distance* [2], yang dapat dilihat pada persamaan sebagai berikut:

$$d(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2}$$

## 2. Pembahasan

Dalam penelitian ini, diharapkan sistem yang di bangun dapat mengelompokkan siswa pada kelas yang ada berdasarkan nilai UN masing-masing siswa. Data siswa yang digunakan adalah data siswa SMA baru yang sudah diterima. Terdapat siswa baru sebanyak 8 orang yang akan di bagi kedalam 3 kelas dengan kapasitas maksimal kelas adalah 3 siswa. Data ini dapat dilihat pada **Tabel 1**.

Seluruh siswa pata pada **Tabel 1** akan di anggap sebagai *cluster* dan akan dilakukan perhitungan jarak/gab. Perhitungan jarak nilai UN dilakukan antar seluruh siswa, hasil perhitungan jarak dapat dilihat pada **Tabel 2**.

**Tabel 1.** Data siswa dan nilai

No.	Nama Siswa	MTK	IPA	Bhs Ing	BHS Ind
1	Budi	7	8	9	4
2	Joko	10	3	6	8
3	Sugi	5	7	3	6
4	Ani	9	5	6	5
5	Nita	7	7	7	6
6	Deva	3	6	6	9
7	Doni	6	5	5	5
8	Ida	8	8	5	6

**Tabel 2.** Hasil Perhitungan Jarak

Jarak Antara	Jarak
Gap(1-2)	= 7.6811457478686
Gap(1-3)	= 6.7082039324994
Gap(1-4)	= 4.7958315233127
Gap(1-5)	= 3
Gap(1-6)	= 7.3484692283495
Gap(1-7)	= 5.1961524227066
Gap(1-8)	= 4.5825756949558
Gap(2-3)	= 7.3484692283495
Gap(2-4)	= 3.7416573867739
Gap(2-5)	= 5.4772255750517
Gap(2-6)	= 7.6811457478686
Gap(2-7)	= 5.4772255750517
Gap(2-8)	= 5.8309518948453
Gap(3-4)	= 5.4772255750517
Gap(3-5)	= 4.4721359549996
Gap(3-6)	= 4.7958315233127
Gap(3-7)	= 3.1622776601684
Gap(3-8)	= 3.7416573867739
Gap(4-5)	= 3.1622776601684
Gap(4-6)	= 7.2801098892805
Gap(4-7)	= 3.1622776601684
Gap(4-8)	= 3.4641016151378
Gap(5-6)	= 5.1961524227066
Gap(5-7)	= 3.1622776601684
Gap(5-8)	= 2.4494897427832
Gap(6-7)	= 5.1961524227066
Gap(6-7)	= 6.2449979983984
Gap(7-8)	= 3.7416573867739

Setelah jarak seluruh antar cluster diketahui maka dapat dilihat kelas yang paling kecil nilai jaraknya yaitu antara kelas ke 5 dan 8. Mereka akan digabungkan menjadi satu cluster kemudian nilai mereka akan di rata-ratakan sehingga menjadi seperti pada Tabel 3.

Tabel 3. Hasil Perhitungan Jarak

No.	Nama Siswa	MTK	IPA	Bhs Ing	BHS Ind
1	Budi	7	8	9	4
2	Joko	10	3	6	8
3	Sugi	5	7	3	6
4-7	Ani,doni	7,5	7,5	6	6
5	Nita	7	7	7	6
6	Deva	3	6	6	9
8	Ida	8	8	5	6

Dalam proses penggabungan cluster ini, jika jumlah kelas telah melebihi kapasitas atau rata2 jumlah siswa dalam satu kelas maka cluster tersebut tidak ikut dalam perhitungan gab karena dianggap telah memenuhi kapasitas kelas. proses di atas akan di ulang sampai jumlah kelas yang terbentuk telah sesuai dengan jumlah yang di inginkan. Hasil akhir dari clustering dapat dilihat pada Tabel 4.

Tabel 3. Hasil Akhir clustering

Kelas	Anggota	MTK	IPA	Bhs Ing	BHS Ind
1	4,5,8	8.25	6.25	6	5.5
2	6,3,7	4.25	6	5	7.25
3	1,2	8.5	5.5	7.5	6

Dalam pembangunan aplikasi diharapkan agar aplikasi dapat menerapkan algoritma AHC untuk mengelompokkan siswa baru seperti perhitungan manualnya. Aplikasi dibangun berbasis web untuk memudahkan dalam pengaksesan data-data yang berasal dari SI Akademik.

Data kelas dan tahun ajar dapat di peroleh dengan mengakses SI Akademik. Sedangkan data siswa beserta nilainya dapat di inputkan oleh panitia penerimaan siswa baru. Dalam pembangunan aplikasi, Tahap-tahap yang dilakukan adalah:

**A. Perencanaan**

Dalam tahap ini dilakukan identifikasi kebutuhan user, fungsional dan nonfungsional untuk mencapai tujuan. Beberapa kebutuhan fungsional yang teridentifikasi adalah:

1. Sistem mampu ngambil data kelas dan tahun ajar dari SI Akademik.
2. Sistem mampu mengolah data siswa bagu dan nilainya.
3. Sistem dapat mengolah data panitia

4. Sistem dapat mengelompokkan siswa pada kelas yang ada berdasarkan nilai UN masing-masing siswa.

Beberapa kebutuhan non fungsional yang teridentifikasi adalah:

1. Dibutuhkan set komputer server sebagai tempat menyimpan aplikasi dan datanya.
2. Dibutuhkan jaringan internet agar server dapat di akses oleh client
3. Dibutuhkan set komputer client untuk menjalankan aplikasi.

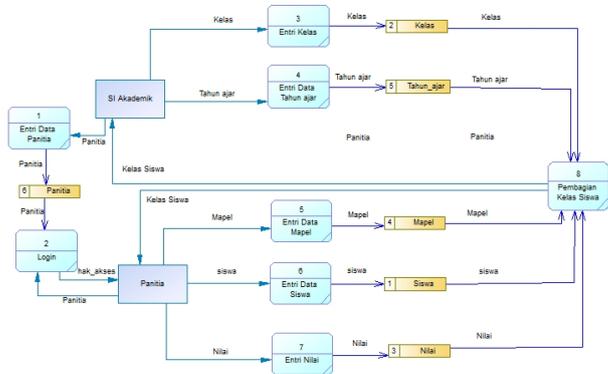
Dalam aplikasi ini dibutuhkan satu jenis user saja yaitu panitia. Panitia dibutuhkan untuk dapat menginputkan data siswa baru beserta nilainya.

**B. Perancangan**

Dalam membuat aplikasi diperlukan perancangan proses dan database untuk mempermudah implementasi. Untuk merancang proses digunakan DFD agar proses-proses dapat tergambar dengan baik. Gambar DFD level 1 dapat dilihat pada Gambar 1 dan DFD level 2 pada Gambar 2.



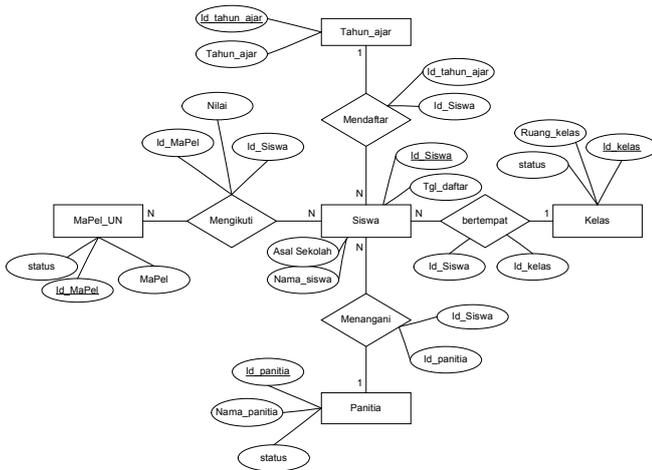
Gambar 1. DFD Level 0



Gambar 2. DFD Level 1

Pada DFD level 1 telah digambarkan bahwa didalam sistem yang akan dibangun adalah terdapat 2 entitas luar yaitu SI Akademik dan Panitia. SI Akademik akan berperan sebagai sumber data kelas, panitia dan tahun ajar. Sedangkan entitas panitia berperan sebagai penginput data mapel, siswa dan nilainya. Dari sumber data tersebut maka akan dilakukan sebanyak 5 proses untuk memasukkan data ke database dan 1 proses untuk menerapkan algoritma AHC dalam mengelompokkan siswa.

Dalam merancang database, digunakan ERD. Gambar ERD dapat di lihat pada **Gambar 3**.

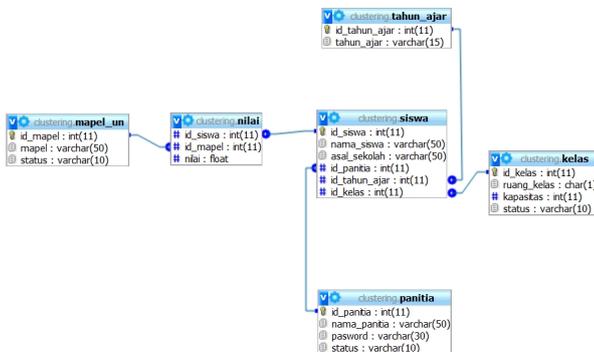


**Gambar 3. ERD**

Dalam rancangan ERD diatas terdapat 4 entitas yang akan dikonversi menjadi tabel, selain itu hubungan banyak ke banyak akan di buat tabel baru yaitu pengbungan antara entitas siswa dengan MaPel\_UN.

**C. Implementasi**

Berdasarkan rancangan ERD, dibuatlah database dengan MySQL pada XAMPP. Tabel di dalam database dan relasinya dapat dilihat pada **Gambar 4**.



**Gambar 4. Tabel dan relasinya**

Aplikasi dibuat berbasis web dengan bahasa pemrograman HTML dan PHP menggunakan Aplikasi server XAMPP. Untuk menerapkan proses algoritma AHC pada proses clustering, pertama-tama dilakukan query data-data dari dalam database, seperti data kelas, mapel, tahun ajar, siswa dan nilainya. Data siswa dan nilainya digabungkan dan disimpan dalam array multidimensi dimana baris adalah mewakili siswa dan kolomnya mewakili nilai mapel sehingga data tersimpan seperti sebuah tabel. Setelah data tersimpan di dalam array multidimensi, maka data dapat ditampilkan seperti pada **Gambar 5**.

Dari array yang ada maka dapat dilakukan perhitungan jarak antar baris atau cluster. Perhitungan dimasukkan dalam perulangan sehingga proses perhitungan akan dapat dilakukan pada semua baris array atau cluser yang ada.

Setelah jarak antar semua cluster diketahui maka dipilih jarak cluster yang paling kecil nilainya. Cluster dengan nilai jarak terkecil akan digabung menjadi baris baru dan nilai dari matkul masing-masing cluster yang digabung akan di rata-ratakan dan menjadi isi dari kolom baris yang baru tersebut.

Setelah baris baru tercipta maka dapat dilakukan penghapusan baris cluster yang telah digabung tadi. Penghapusan dilakukan dengan membuat array baru yang berisi data dari baris yang seakan-akan telah dihapus. Setelah dilakukan penggabungan dan penghapusan data dapat ditampilkan seperti pada **Gambar 6**. Array baru ini akan digunakan sebagai pembanding pada perulangan perhitungan jarak selanjutnya.

Perhitungan jarak dilakukan berulang-ulang sampai jumlah cluster yang masih ada sesuai dengan yang diinginkan yaitu rata-rata kapasitas kelas yang ada. Pada perulangan yang kedua dapat dilihat hasilnya pada **Gambar 7**, karena cluster yang baru berasal dari penggabungan dari tiga siswa, yaitu mencapai jumlah max kapasitas kelas. maka cluster tersebut tidak dihitung lagi jaraknya dengan cluster yang lain. Proses perulangan ke tiga dan empat dapat dilihat pada **Gambar 8** dan **Gambar 9**.

Pada perulangan ke lima, jumlah cluster atau kelas telah sesuai dengan yang di inginkan sehingga perulangan berhenti dan hasilnya dapat di tampilkan. Hasil akhir dari clustering dapat dilihat pada **Gambar 10**.

	0	1	2	3	anggota
0	7	8	9	4	0
1	10	3	6	8	1
2	5	7	3	6	2
3	9	5	6	5	3
4	7	7	7	6	4
5	3	6	6	9	5
6	6	5	5	5	6
7	8	8	5	6	7

**Gambar 5. Tampilan data siswa di aplikasi**

Cluster 4-7 digabung menjadi index ke 8 = | 7.5 | 7.5 | 6 | 6

	0	1	2	3	anggota
0	7	8	9	4	0
1	10	3	6	8	1
2	5	7	3	6	2
3	9	5	6	5	3
5	3	6	6	9	5
6	6	5	5	5	6
8	7.5	7.5	6	6	4-7

Gambar 6. Perulangan pertama perhitungan jarak

Cluster 3-8 digabung menjadi index ke 9 = | 8.25 | 6.25 | 6 | 5.5

	0	1	2	3	anggota
0	7	8	9	4	0
1	10	3	6	8	1
2	5	7	3	6	2
5	3	6	6	9	5
6	6	5	5	5	6
9	8.25	6.25	6	5.5	3-4-7

Gambar 7. Perulangan kedua perhitungan jarak

Cluster 2-6 digabung menjadi index ke 10 = | 5.5 | 6 | 4 | 5.5

	0	1	2	3	anggota
0	7	8	9	4	0
1	10	3	6	8	1
5	3	6	6	9	5
9	8.25	6.25	6	5.5	3-4-7
10	5.5	6	4	5.5	2-6

Gambar 8. Perulangan ketiga perhitungan jarak

Cluster 5-10 digabung menjadi index ke 11 = | 4.25 | 6 | 5 | 7.25

	0	1	2	3	anggota
0	7	8	9	4	0
1	10	3	6	8	1
9	8.25	6.25	6	5.5	3-4-7
11	4.25	6	5	7.25	5-2-6

Gambar 9. Perulangan keempat perhitungan jarak

Cluster 0-1 digabung menjadi index ke 12 = | 8.5 | 5.5 | 7.5 | 6

	0	1	2	3	anggota
9	8.25	6.25	6	5.5	3-4-7
11	4.25	6	5	7.25	5-2-6
12	8.5	5.5	7.5	6	0-1

Anggota kelas ke 1 = siswa dengan index ke 3-4-7  
 Anggota kelas ke 2 = siswa dengan index ke 5-2-6  
 Anggota kelas ke 3 = siswa dengan index ke 0-1

Gambar 10. Perulangan kelima perhitungan jarak

Perhitungan di ulang sebanyak 5 kali untuk mendapatkan jumlah cluster yang di inginkan.

### 3. Kesimpulan

Hasil perhitungan di dalam program aplikasi telah sesuai dengan perhitungan manual, berarti program telah berjalan dengan baik. Dalam perhitungan di dalam aplikasi, nilai-nilai dari perhitungan dapat disimpan sementara di dalam array multidimensi untuk dapat digunakan sebagai bahan perhitungan selanjutnya.

### Daftar Pustaka

- [1] [http://www.ub.edu/included/docs/INCLUDED\\_actions%20for%20success.pdf](http://www.ub.edu/included/docs/INCLUDED_actions%20for%20success.pdf): "Actions For Succes in School in Europe", 17 Februari 2013.
- [2] J. Han and K. Micheline, *Data Mining Concepts and Technique*, Morgan Kaufmann, 2001.
- [3] R. Margo dan E. Ari Eka, *penerapan algoritma clustering untuk mengelompokkan ketertarikan siswa berdasarkan aktivitas di model pembelajaran elektronik*, 2014.
- [4] T. Humasak, *implementasi metode hierarchical clustering pada data genetic microarray*. 2014.

### Biodata Penulis

**Marjiyono**, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Sistem Informatika STMIK AMIKOM Yogyakarta, lulus tahun 2011. Memperoleh gelar. Saat ini sedang menempuh study lanjut di Magister Tehnik Informasi STMIK AMIKOM YOGYAKARTA

